size, author, etc.). The domain expert is then asked to construct the predominate semantic structure of the database by identifying the primary conceptual groupings that are repeatedly used through out the descriptions. The domain expert is also asked to assign each conceptual grouping an importance (high, medium, low or none) as it relates to the content of a description. [For example, the brand is more important in a description of a bicycle than its color is.] These groupings and their importance are recorded in the Lexicon file.

Figure 4 illustrates how the core vocabulary is supplemented and associated within the conceptual groupings that form the semantic structure. The computer program generates a random sample of descriptions from the database for each term in the core vocabulary developed in Figure 2 that is representative of the population at a 95% confidence level. The citations for each term are presented to the domain expert along with the list of primary conceptual groupings developed in Figure 3. The domain expert is asked to assign each term to a primary conceptual grouping. The computer program then records all of the terms and their conceptual grouping assignments in the Lexicon file. The computer program then prepares a listing of all core vocabulary terms within each conceptual grouping. The listing is presented to the domain expert who is requested to identify any additional terms that are appropriate to each conceptual grouping, including synonyms and common misnomers [i.e. "dungarees" and "jeans" to the group of "clothing types"]. These additional terms are recorded in the Lexicon file with their conceptual grouping assignments.

Figure 5 illustrates how the invention captures the contextual significance of the usage of each term. The computer program prepares a record for each term that starts with it as the records "primary term" and then lists all of the other terms in the Lexicon file that have the same conceptual grouping assignment. The domain expert is then presented with the primary term and its associated terms and asked to identify each associated term's relationship to the primary term [i.e. synonym, misnomer, similar term, no relationship, anonym]. These contextual relationships are recorded in the Lexicon file. The computer program then determines a significance factor for each term in each record based on the importance of the conceptual grouping and the relationship of the term in context to the primary term. These factors are stored in a two-dimensional matrix "look up" table.

Figure 6 illustrates how the invention interrupts the meaning of the user request. The user enters one or more words that describe the entries they are interested in. The computer program parses the input into individual query terms and assigns each a significance factor of 1.0. The computer program then compares each query term with each primary term in the Lexicon file using a character string matching function. When an exact match is found, the significance factor of the inputted query term is reset to the value of the primary term in the Lexicon file. All terms associated with the primary term are then added to the list of query terms along with their significance factors. This process is repeated for every query term from the user request. When complete, the set of query terms and their significance factors represent the meaning of the user request in the semantic structure of the database.

Figure 7 illustrates how the invention determines the similarity of the content a database record and the meaning of a user request. The computer program creates a similarity index for each record in the database and sets all of them to 0.0. The computer program then takes each query term and executes a character string comparison with each word in the first database description. If there is an exact match, the query term's significance factor is added to the database record's similarity index. If an exact match is not found, no change is made to the database record's similarity index. The process is repeated with the next query term until all query terms have been compared to the database record's description. When all query terms have been compared with the database record description, the computer program repeats the entire procedure on the next database record. In this manner, the similarity between the content of each database record and the meaning of the user request is captured in a quantative index. The significance factors developed in Figure 6 were designed so that high values of the similarity index represent close matches and negative values indicate that database record and the meaning of the user request are dissimilar in a meaningful way. [i.e. if the user requested "plate", "platter" would have a high similarity index but "bowl" would have a negative value]. The computer program then sorts the records with positive similarity indexes in descending order for presentation for subjective review by the user.

- 9 -